	Type	L #	Hits	Search Text	DBs	Time Stamp
1	BRS	L2	25	("6418433" or "6356899" or "6334131" or "6285999" or "6282511" or "6125361" or "6112203" or "5895470" or "5855015" or "5835905" or "6067539" or "6353827" or "6507841").pn.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:07
2	BRS	L3	13	"1" and (hypertext or hyper-text or (hyper adj text)) and page	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:27
3	BRS	L4	0	3 and (relevancy near3 page) with (keyword or key-word or (key adj word))	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:44
4	BRS	L5	0	2 and (relevancy near3 page) with (keyword or key-word or (key adj word))	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:10
5	BRS	L6	8	(relevancy near3 page) with (keyword or key-word or (key adj word))	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:11

	Type	L #	Hits	Search Text	DBs	Time Stamp
6	BRS	L7	10	(relevancy near3 page) same (keyword or key-word or (key adj word))	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:12
7	BRS	L8	20	(relevancy near3 page) and (keyword or key-word or (key adj word))	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:14
8	BRS	L9	4	3 and (relevan$4 with page) and (keyword or key-word or (key adj word))	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:16
9	BRS	L10	13	2 and (hypertext or hyper-text or (hyper adj text)) and page	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:15
10	BRS	L11	2	10 and (relevan$4 same page) same (keyword or key-word or (key adj word))	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:17

	Type	L #	Hits	Search Text	DBs	Time Stamp
11	BRS	L12	5	10 and (relevan$4 same page) and (keyword or key-word or (key adj word))	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:18
12	BRS	L13	3	2 and weight$4 near page	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:19
13	BRS	L14	79	page adj weight$4	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:22
14	BRS	L15	2	14 and content adj scor$4	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:24
15	BRS	L16	97	content adj scor$4	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:25

	Type	L #	Hits	Search Text	DBs	Time Stamp
16	BRS	L17	15	16 and (707/$.ccls. or 709/$.ccls. or 345/$.ccls. or 704/$.ccls or 706/$.ccls. or 715/$.ccls.)	USPAT; US-PG PUB; EPO; JPO; DERW ENT; IBM_TD B	2003/04/04 11:26
17	BRS	L18	8	17 and page	USPAT; US-PG PUB; EPO; JPO; DERW ENT; IBM_TD B	2003/04/04 11:27
18	BRS	L19	29	16 and page	USPAT; US-PG PUB; EPO; JPO; DERW ENT; IBM_TD B	2003/04/04 11:27
19	BRS	L20	6	2 and (hypertext or hyper-text or (hyper adj text)) and page and weight$4 and scor$4 and content and relevan$4	USPAT; US-PG PUB; EPO; JPO; DERW ENT; IBM_TD B	2003/04/04 11:34
20	BRS	L22	2	20 and function and frequenc$4	USPAT; US-PG PUB; EPO; JPO; DERW ENT; IBM_TD B	2003/04/04 11:30

	Type	L #	Hits	Search Text	DBs	Time Stamp
21	BRS	L21	6	20 and (computing or compute! or computed! or computes! or calculat$6)	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:31
22	BRS	L23	5	2 and (hypertext or hyper-text or (hyper adj text)) and page and weight$4 and scor$4 and content and relevan$4 and rank$4	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:43
23	BRS	L24	199	(hypertext or hyper-text or (hyper adj text)) and page and weight$4 and scor$4 and content and relevan$4 and rank$4	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:44
24	BRS	L25	3	24 and (relevancy near6 page) and (keyword or key-word or (key adj word))	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:46
25	BRS	L26	26	24 and (rank$4 near4 page) and (keyword or key-word or (key adj word))	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:47

	Type	L #	Hits	Search Text	DBs	Time Stamp
26	BRS	L27	4	26 and frequenc$4 with page with (keyword or key-word or (key adj word))	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:48
27	BRS	L28	7	26 and frequenc$4 with (keyword or key-word or (key adj word))	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:50
28	BRS	L29	13	2 and access	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:50
29	BRS	L30	14	2 and access$5	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 11:50
30	BRS	L31	15	2 and (graph or connect$6)	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 12:00

	Type	L #	Hits	Search Text	DBs	Time Stamp
1	BRS	L1	456	chakrabarti.in.	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 09:44
2	BRS	L2	2	1 and alam.xp.xa.	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 09:46
3	BRS	L4	0	1 and alam.xa.	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 09:46
4	BRS	L5	28	1 and frequenc$5	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 09:46
5	BRS	L3	1	1 and alam.xp.	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 09:49

	Type	L #	Hits	Search Text	DBs	Time Stamp
6	BRS	L6	5	5 and page and weight$4 and scor$4	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 10:04
7	BRS	L8	1	6 and anchor	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 09:54
8	BRS	L7	1	6 and anchor adj text	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 09:55
9	BRS	L9	0	1 and font	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 09:55
10	BRS	L10	0	1 and sizet	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 09:55

	Type	L #	Hits	Search Text	DBs	Time Stamp
11	BRS	L11	91	1 and size	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 09:56
12	BRS	L12	5	1 and page and weight$4 and scor$4 and web and hypertext and link$4 and hyperlink$4	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 10:09
13	BRS	L13	1	12 and visit$4	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 10:06
14	BRS	L14	3	12 and probability	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 10:08
15	BRS	L15	3481	probability with weight$4	USPAT; US-PG PUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 10:09

	Type	L #	Hits	Search Text	DBs	Time Stamp
16	BRS	L16	11	15 and page and weight$4 and scor$4 and web and hypertext and link$4 and hyperlink$4	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/04/04 10:09

Patent Assignment Abstract of Title

Total Assignments: 1

Application #: 09757435 **Filing Dt:** 01/10/2001 **Patent #:** NONE **Issue Dt:**

PCT #: NONE **Publication #:** 20020129014 **Pub Dt:** 09/12/2

Inventors: Brian S. Kim, Sudong Chung, Anurag Dod, Michael Kim, Yeogirl Yun

Title: Systems and methods of retrieving relevant information

Assignment: 1

Reel/Frame:	**Received:**	**Recorded:**	**Mailed:**	**Page**
011525/0801	02/23/2001	02/12/2001	05/02/2001	

Conveyance: ASSIGNMENT OF ASSIGNORS INTEREST (SEE DOCUMENT FOR DETAILS).

Assignors:

KIM, BRIAN S.	**Exec Dt:** 02/06/2001
CHUNG, SUDONG	**Exec Dt:** 02/06/2001
DOD, ANURAG	**Exec Dt:** 02/06/2001
KIM, MICHAEL	**Exec Dt:** 02/06/2001
YUN, YEOGIRL	**Exec Dt:** 02/06/2001

Assignee: WISENUT, INC. A CORPORATION OF THE STATE OF DELAWARE
SUITE 269
1333 LAWRENCE EXPRESSWAY
SANTA CLARA, CALIFORNIA 95051

Correspondent: ROBERT MOLL
1173 SAINT CHARLES COURT
LOS ALTOS, CA 94024

Search Results as of: 4/4/2003 9:51:19 A.M.

US-PAT-NO: **5835905**

DOCUMENT-IDENTIFIER: US 5835905 A

TITLE: System for predicting documents relevant to focus documents by spreading activation through network representations of a linked collection of documents

--------- KWIC --------

5835905

The currently preferred embodiment of the present invention is implemented for analyzing collections of linked documents residing on the portion of the Internet known as the World Wide Web (hereinafter the Web). However, it should
be noted that the present invention is not limited to use on the Web and may be utilized in any system which provides **access** to linked entities, including documents, images, videos, audio, etc.. The following terms defined herein are familiar to users of the Web and take on these familiar meanings:

World-Wide Web or Web: The portion of the Internet that is used to store and **access** linked documents.

Web Page or Page: A document **accessible** on the Web. A Page may have multi-media content as well as relative and absolute links to other pages.

The basic steps for categorizing web pages in a web locality and for predicting relevance of other pages of a selected page as may be performed in the currently preferred embodiment of the present invention are briefly described with reference to the flowchart in FIG. 1. First, raw data is gathered for the web locality, step 101. Such raw data may be obtained from usage records or **access** logs of the web locality and by direct traversal of the Web pages in the Web locality. As described below, "Agents" are used to collect such raw data. However, it should be noted that the described agents are not the only possible method for obtaining the raw data for the basic feature vectors. It is anticipated that Internet service providers have the capabilities to provide such raw data and may do so in the future.

In any event, the raw data is then processed into desired formats for

performing the categorization (feature vectors) and relevance prediction (topology, usage path and text similarity maps), step 102. The raw data is comprised of topology information, page meta-information, page frequency path information and text similarity information. Topology information describes the hyperlink structure among Web pages at a Web locality. Page meta-information defines various features of the pages, such as file size and URL. ~~Usage frequency and path information indicate how many times a Web page~~ has been **accessed** and how many times a traversal was made from one Web ~~page to~~ another. Text similarity information provides an indication of the similarity of text among all text Web pages at a Web locality.

Usage frequency and usage paths, which indicate how many times a Web page has been **accessed** and how many times a traversal was made from one Web ~~page to~~ another.

The site's topology is ascertained via "the walker", an autonomous agent that, given a starting point, performs an exhaustive breadth-first traversal of pages within the Web locality. FIG. 2 is a flowchart illustrating the steps performed by the walker. Referring to FIG. 2, the walker uses the Hypertext Transfer Protocol (HTTP) to request and retrieve a web page, step 201. The walker may also be able to **access** the pages from the local filesystem, bypassing the HTTP. The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve, step 203. The meta-information for the page is also extracted and stored, step 204. The meta-information includes at least the following page meta-information: name, title, list of children (pages associated by hyperlinks), file size, and the time the page was last modified. The page is then added to a topology matrix, step 205. The topology matrix represents the page to page hypertext relations, and a set of meta-information called the meta-document vectors, which represents the meta-information for each Web page The list of pages to request and retrieve is then used to obtain the next page, step 206. The process then repeats per step 202 until all of the pages on the list have been retrieved.

Thus, the walker produces a graph representation of the hyperlink structure of the Web locality, with each node having at least the above described meta-information. It is salient to note that the walker may not have reached all nodes that are **accessible** via a particular server--only those nodes that were reachable from the starting point (e.g. a Home Page for the Web locality) are included. This can be alleviated by walking the local filesystem the

locality resides on.

Most servers have the ability to record transactional information, i.e. **access** logs, about requested items. This information usually consists of at least the time and the name of the URL being requested as well as the machine name making the request. The latter field may represent only one user making requests from their local machine or it could represent a number of users whose requests are being issued through one machine, as is the case with firewalls and proxies. This makes differentiating the paths traversed by individual users from these **access** logs non-trivial, since numerous requests from proxied and firewalled domains can occur simultaneously. That is, if 200 users from behind a proxy are simultaneously navigating the pages within a site, how does one determine which users took which paths? This problem is further complicated
by local caches maintained by each browser and intentional reloading of pages by the user.

The technique implemented to determine user's paths, a.k.a. "the whittier", utilizes the Web locality's topology along with several heuristics. FIG. 3 is a flowchart illustrating the steps performed to determine user paths. First, a user path is obtained from the web locality **access** logs, step 301. The topology matrix is consulted to determine legitimate traversals. It is then determined if there are any ambiguities with respect to the user path, step 302. As described above such ambiguities may arise in the situation where the request is from a proxied or firewalled domain. If an ambiguity is suspected, predetermined heuristics are used to disambiguate user paths, step 303. The heuristics used relies upon a least recently used bin packing strategy and session length time-outs as determined empirically from end-user navigation patterns. Essentially, new paths are created for a machine name when the time between the last request and the current request was greater than the session boundary limit, i.e., the session timed out. New paths are also created when the requested page is not connected to the last page in the currently maintained path. These tests are performed on all paths being maintained for that machine name, with the ordering of tests being the paths least recently extended. The foregoing analysis produces a set of paths requested by each machine and the times for each request.

To the extent that the properties that help users navigate around the space and remember locations or ones that support the unit tasks of the user's work, the visualizations provide value to the user. Visualizations can be applied to the Web by treating the pages of the Web as objects with properties. Each of these visualizations provide an overview of a Web locality in terms of some simple property of the pages. For example, the present invention may be used in support of information visualization techniques, such as the WebBook

described in co-pending and commonly assigned application Ser. No. 08/525,936
entitled "Display System For Displaying Lists of Linked Documents" now pending,
to form and present larger aggregates of related Web pages. Other examples include a Cone Tree which shows the connectivity structure between pages and a
Perspective Wall which shows time-indexed **accesses** of the pages. The cone tree
is described in U.S. Pat. No. 5,295,243 entitled "Display of Hierarchical Three-Dimensional Structures With Rotating Substructures". The Perspective Wall is described in U.S. Pat. No. 5,339,390 entitled "Operating A Processor To Display Stretched Continuation Of A Workspace". Thus, these visualizations are based on one or a few characteristics of the pages.

The computer based system on which the currently preferred embodiment of the
present invention may be implemented is described with reference to FIG. 14. The computer based system and associated operating instructions (e.g. software)
embody circuitry used to implement the present invention. Referring to FIG. 14, the computer based system is comprised of a plurality of components coupled
via a bus 1401. The bus 1401 may consist of a plurality of parallel buses (e.g. address, data and status buses) as well as a hierarchy of buses (e.g. a processor bus, a local bus and an I/O bus). In any event, the computer system is further comprised of a processor 1402 for executing instructions provided via bus 1401 from Internal memory 1403 (note that the Internal memory 1403 is typically a combination of Random **Access** and Read Only Memories). The processor 1402 will be used to perform various operations in support extracting raw data from web localities, converting the raw data into the desired feature vectors and topology, usage path and text similarity matrices, categorization and spreading activation. Instructions for performing such operations are retrieved from Internal memory 1403. Such operations that would be performed by the processor 1402 would include the processing steps described in FIGS. 1-4
and 7. The operations would typically be provided in the form of coded instructions in a suitable programming language using well-known programming techniques. The processor 1402 and Internal memory 1403 may be discrete components or a single integrated device such as an Application Specification Integrated Circuit (ASIC) chip.

9. The method as recited in claim 8 wherein said step of obtaining raw data for said linked collection of documents further comprising the steps of

obtaining **access** data from said linked collection, said **access** data indicating when and from where documents in said linked collections have been **accessed**.

c2) generating topology characteristic information and usage path characteristic information from said raw data, said topology information for indicating if a document contains a link to another document and said usage path information indicating the number of times a document was **accessed** from another document; and